

Published in final edited form as:

J Proteome Res. 2009 August ; 8(8): 3944–3950. doi:10.1021/pr900251d.

Systematical Optimization of Reverse-phase Chromatography for Shotgun Proteomics

Ping Xu¹, Duc M. Duong^{1,2}, and Junmin Peng^{1,2,3}

¹ Department of Human Genetics, Center for Neurodegenerative Diseases, Emory University, Atlanta, GA 30322

² Emory Proteomics Service Center, Emory University, Atlanta, GA 30322

Summary

We report the optimization of a common LC/MS/MS platform to maximize the number of proteins identified from a complex biological sample. The platform uses digested yeast lysate on a 75 μm internal diameter \times 12 cm reverse-phase column that is combined with an LTQ-Orbitrap mass spectrometer. We first generated a yeast peptide mix that was quantified by multiple methods including the strategy of stable isotope labeling with amino acids in cell culture (SILAC). The peptide mix was analyzed on a highly reproducible, automated nanoLC/MS/MS system with systematic adjustment of loading amount, flow rate, elution gradient range and length. Interestingly, the column was found to be almost saturated by loading $\sim 1 \mu\text{g}$ of the sample. Whereas the optimal flow rate ($\sim 0.2 \mu\text{l}/\text{min}$) and elution buffer range (13–32% of acetonitrile) appeared to be independent of the loading amount, the best gradient length varied according to the amount of samples: 160 min for 1 μg of the peptide mix, but 40 min for 10 ng of the same sample. The effect of these parameters on elution peptide peak width is evaluated. After full optimization, 1,012 proteins (clustered in 806 groups) with an estimated protein false discovery rate of $\sim 3\%$ were identified in 1 μg of yeast lysate in a single 160-min LC/MS/MS run.

Keywords

liquid chromatography; LTQ-Orbitrap; shotgun proteomics; loading amount; flow rate; gradient range; gradient length

Introduction

In the last decade, mass spectrometry has emerged as a central proteomics technology in the post-genomic era. Shotgun (bottom-up) proteomics is the most commonly used platform for analyzing proteins and posttranslational modifications^{1–3}. In a typical protocol, simple or complex protein samples are digested by proteases (e.g. trypsin) to generate peptides that are further analyzed by reverse-phase liquid chromatography coupled with tandem mass spectrometry (LC/MS/MS). The MS/MS spectra are then searched against protein databases, resulting in protein identification and determination of post-translational modification sites. Additional strategies such as label-free and stable isotope labeling methods are implemented to obtain quantitative data⁴. Despite rapid development, current LC/MS/MS platforms still lack the sensitivity and throughput to detect all proteins from mammalian cells in a single experiment. To achieve successful analyses of complex protein samples, it is important to

³Address correspondence to Junmin Peng, Ph: 404.712.8510; Fax: 404.727.3728; E-mail: jpeng@emory.edu.

maximize protein/peptide analytic power by optimizing liquid chromatography and mass spectrometry settings.

Liquid chromatography of peptides prior to MS is usually achieved on a reverse-phase column which offers high-resolution separation capacity and utilizes mobile phase solvents compatible with electrospray ionization. As LC efficiency increases with smaller internal dimension and longer columns, detection sensitivity is greatly improved with the development of online microcapillary LC with internal dimension less than 150 μm . A common LC platform includes a standard HPLC, a flow splitter, and a 75 μm I.D. \times 12 cm reverse-phase column³. Further decreasing the column I.D. and increasing the column length are possible⁴, but “ultra-high-pressure” LC systems would be required to provide sufficient back pressure for solvent delivery at optimum column flow rates⁵. In addition, chromatography peak capacity is also influenced by LC elution gradients and analysis time^{6–8}.

Recently, the development of LTQ-Orbitrap hybrid mass spectrometer offers high-resolution precursor ion scans in the Orbitrap and sensitive, rapid acquisition of MS/MS scans in the LTQ. Compared to a 3D ion trap, the LTQ confines ions in a 2D radiofrequency field, providing higher storage capacity, faster scan rate and better detection efficiency⁹. The Orbitrap captures ions by orbital trapping, with electrostatic fields generated by central and outer electrodes^{10, 11}. The ions move in spirals around the central electrode and oscillate along the z -axis. The axial oscillation of the ions is independent of initial energy, directions and positions, and is recorded as a current image. The image is then converted to ion frequencies by Fourier transform, leading to highly accurate measurements of m/z values in a large dynamic range¹². By combining the advantages of both LTQ and Orbitrap, the hybrid instrument has been demonstrated to be a powerful tool for proteomic studies^{13–16}.

Although microcapillary LC parameters have been extensively studied with respect to peak capacity, there is no detailed report on how to adjust sample loading and LC parameters to optimize protein identification, especially in the context of complex mixtures using the recently developed LTQ-Orbitrap mass spectrometer. Here we used a step-wise protocol to perform a series of optimization on numerous parameters, which are described in a shotgun proteomic study using a complex biology sample (i.e. yeast lysate) among more than 50 LC/MS/MS runs.

Materials and Methods

Sample Preparation from Yeast *S. cerevisiae*

A yeast strain SUB592¹⁷ was grown in YPD medium at 30°C to early log phase ($A_{600} = 1.0$) and extracted in lysis buffer (10 mM Tris-HCl, pH 8.0, 0.1 M NaH_2PO_4 , 8 M urea, 0.02% SDS and 10 mM β -mercaptoethanol). The SILAC analysis was performed in a similar protocol as described¹⁸. The isogenic yeast strain JMP025 was generated with *lys2* and *arg4* gene deletions. The strain was grown in heavy synthetic medium (0.7% Difco yeast nitrogen base, 2% dextrose, supplemented with adenine, uracil, and amino acids plus 12 mg/L [$^{13}\text{C}_6$, $^{15}\text{N}_4$] Arg and 18 mg/L [$^{13}\text{C}_6$] Lys (Cambridge Isotope Laboratories, Andover, MA) for >8 generations until A_{600} was ~ 0.7 . The cells were then harvested and lysed in the same lysis buffer.

Protein Quantification

Protein concentration of yeast lysate was measured by a standard BCA protein assay kit (Thermo Scientific, Rockford, IL) and by a Coomassie stained SDS gel. In the gel analysis, proteins were concentrated on a very short 9% SDS gel (~ 2 mm long), stained with Coomassie Blue G250, and quantified by Scion Image (<http://rsb.info.nih.gov/nih-image/>). In both methods, bovine serum albumin (BSA) was used as standard.

Protein Digestion and Peptide Purification

The lysate (2 mg protein) was reduced with 10 mM DTT at 37°C for 30 min and alkylated with 50 mM iodoacetamide (IAA) in the dark at room temperature for 30 min. The sample was then diluted to 2 M urea with buffer (5% AcN in 50 mM NH₄HCO₃), and digested with trypsin (40 µg) at 37°C overnight. The resulting peptide solution was cleaned with a Vydac Bioselect 218 SPE1000 C₁₈ cartridge (Chrom Tech, Apple Valley, MN), dried and dissolved with sample loading buffer (6% acetic acid, 0.005% heptafluorobutyric acid [HFBA], 0.1% TFA, and 5% AcN). The SILAC-labeled sample was processed under the same conditions.

Protein Identification by LC/MS/MS

A hybrid LTQ-Orbitrap MS (Thermo Scientific) equipped with an Agilent 1100 binary HPLC (Agilent Technologies, Palo Alto, CA), a Famos autosampler (LC Packings, San Francisco, CA), and a 75 µm I.D. × 12 cm fused-silica capillary column was applied for all runs. The column was packed with C₁₈ resins (5 µm magic C₁₈AQ; pore size, 200 Å; Michrom Bioresources, Auburn, CA). Column flow rate was measured by calibrated 5 µl micropipets for at least 3 times (VWR, West Chester, PA). Peptide samples were loaded onto the column by the autosampler, and eluted by a designed gradient (buffer A, 0.4% acetic acid, 0.005% HFBA, and 5% AcN; buffer B, 0.4% acetic acid, 0.005% HFBA, and 95% AcN). The eluted peptides were detected in a precursor MS scan by Orbitrap (400–1600 *m/z*, 60,000 resolution at *m/z* 400, 1 µscan, and 1 × 10⁶ for automatic gain control), followed by sequential data-dependent MS/MS scans of the ten most abundant ions (minimal ion intensity of 500 counts, isolation width of 2 *m/z*, 35% normalized collision energy, 1 µscan, target value of 5,000 for automatic gain control, 60 sec dynamic exclusion, preview mode enabled, removal of 1+ ions or ions with unassigned charge state, and selection of 2+, 3+, and 4+ ions). When the LTQ was used as survey scan MS analyzer, all of parameters were the same with the exclusion of high-resolution, preview mode function, and charge state selection.

Database Search

The MS/MS spectra were searched by the Sequest-Sorcerer algorithm on a Sorcerer 2 IDA (Sage-N-Research)¹⁹ against a composite target/decoy database to estimate false discovery rate²⁰. The target proteins included yeast proteins (from www.stanford.edu/saccharomyces) and common contaminants, such as porcine trypsin and human keratins. The decoy proteins were generated from pseudo-reversed sequences of all target proteins²¹. Searching parameters consisted of semi-tryptic restriction, fixed modification of Cys (+57.0215 Da, alkylation by iodoacetamide), and dynamic modification of oxidized Met (+15.9949 Da). Mass tolerance was set to ±20 ppm. For SILAC analysis, dynamic modifications of Arg (+10.0083 Da) and Lys (+6.0201 Da) were included. Only *b* and *y* ions were considered during the database match.

Peptide matches were filtered by a minimal peptide length of 6 amino acids²¹, then grouped by trypticity (only accept fully and partially tryptic peptides) and charge states²⁰. In each group, the peptide matches were further filtered by dynamically increasing XCorr and ΔCn cutoffs until the global protein false discovery rate was ~3%⁷. While effectively removing false matches, the procedure recovered the vast majority (93.2 ± 2.6%) of estimated true MS/MS matches (also named spectral counts, Table 1). The filtering procedure also led to consistent results from technical replicates (Table 1).

When matching filtered peptides to proteins, we assigned the proteins sharing the same peptide (s) in one group, in which the top protein with highest peptide matches was selected to represent the group. For simplicity, we used all identified protein number for comparison during the optimization of LC settings. After optimization, both the identified proteins and protein groups were reported for the analysis of total yeast cell lysate by the LC/MS/MS run. Some of the accepted peptides and spectra are attached (see supplemental Table S3 and S4).

Determination of Peptide Recovery by SILAC

To estimate the peptide recovery for C₁₈ cartridge cleanup, 0.5% of the input and elution was taken and mixed with equal amount of SILAC-labeled peptides (derived from 1 µg of total protein), respectively. The two samples were analyzed by LC/MS/MS and the SILAC-labeled peptides were used as internal standards to evaluate peptide recovery. The detail quantification methods are described in another paper¹⁸.

Results and Discussion

Preparation of an Accurately Quantified Peptide Mixture

We used a highly complex biological sample from yeast to perform the optimization study (Figure 1A). Total protein was extracted from yeast cells using urea and SDS, and quantified by two independent methods. First, we used standard BCA assay in which Cu²⁺ is reduced to Cu¹⁺ by proteins in an alkaline medium and the reduced Cu¹⁺ selectively forms an intense purple complex with bicinchoninic acid to allow colorimetric quantification²². In six repeated analyses, the detected concentration was 4.2 ± 0.1 µg/µl using BSA as standard. To account for possible interference from buffer chemicals, we measured the protein concentration again based on Coomassie-stained SDS gel images (Figure 1B). As the interfering chemicals are removed after gel electrophoresis, the Coomassie dye only interacts with positively charged residues in proteins²³. To minimize quantification errors, we ran a short gel to compress all proteins in 2-mm range. The dye absorbance signal was linear to titrated BSA concentration ($R^2 = 0.986$) in all three replicates (Figure 1B), and the protein concentration of the cell lysate was measured to be 4.5 ± 0.1 µg/µl, consistent with the BCA result. Finally, the averaged concentration (4.35 µg/µl) was used for subsequent assays.

Yeast proteins (2.0 mg) were then reduced, alkylated, digested in solution, and desalted by a C₁₈ cartridge. To evaluate peptide recovery during the desalting step, we used SILAC-labeled heavy peptides as internal standard to quantify >100 abundant peptides in the input and eluate²⁴. For instance, one peptide (NVPLYQHLADLSK) had a relative intensity of 1.2 before desalting and 1.1 after desalting when compared to the heavy standard (Figure 1C). Thus, the recovery of this peptide was ~91.7% (1.1/1.2). According to the recovery rate of 116 different peptides, we calculated final mean value of peptide recovery ($73.7 \pm 15\%$, Figure 1D) and used it to estimate total peptide amount for LC loading.

Evaluation of experimental variation of the LC/MS/MS system

Since reproducibility of the LC system is a prerequisite for reliable comparison of different runs with varying parameters, we tested run-to-run variation by repeated LC/MS/MS runs. A peptide mixture (equivalent to 1 µg of yeast lysate) was analyzed four times on a 75 µm I.D. × 12 cm reverse-phase column using the same parameter settings. Base peak profiles for the replicates were almost identical (Figure 2A) and retention time shifts of the same peptide ions were usually less than 1 min. After database search and filtering, the four runs resulted in highly consistent number of accepted spectra counts, peptides and proteins, with relative standard deviation of 2.8%, 2.4% and 1.5%, respectively (Figure 2B). The data strongly support high reproducibility of the automated LC/MS/MS system used in this study. The same reverse-phase column was used for entire optimization process and column degeneration was not observed after more than 200 runs (data not shown).

Optimization of LC Parameters

First, we examined the effect of peptide loading amount on protein identification (Figure 3A). When peptide samples were titrated from 10 ng to 1 µg on the column, identified protein number was increased from 395 to 699. Further addition of loading amount to 4 µg resulted in only 6%

increase of identified proteins. The titration curve suggests that the LC/MS/MS system was saturated around the point of 1 μg . Similar results were obtained by analyzing the accepted spectral counts and peptide numbers (supplemental Table S1). Whereas loading higher amount of peptides could raise ion intensity, it also led to ion peak broadening that may suppress adjacent co-eluting ions. In the example of an abundant peptide of the TEF2 protein (IGGIGTVPVGR) in the 10 ng run versus the 4 μg (400-fold more loading) run, the ion signal was increased ~ 200 -fold, and the peak width at half height was broadened from 0.22 min to 0.45 min (Figure 3B). The loading saturation of the column was indicated by retention time shift from 40.0 min in the 10 ng run to 27.7 min in the 4 μg run, because the peptide may be pushed forward on the column due to competitive binding of more hydrophobic peptides in the 4 μg run. In addition to this abundant peptide, we analyzed the peak width distribution of all accepted peptides and found the majority of the data could be roughly fitted into Gaussian curve (Figure 3C). The mean values of the Gaussian clearly indicates a global shift of peak width from 0.12 min (10 ng loading) to 0.18 min (4 μg loading), suggesting the occurrence of peak broadening. It should be mentioned that most of peptide peaks were narrower than the abundant TEF2 peptide (Figure 3B). During this analysis, the amount of 1 μg peptides on the 75 μm I.D. \times 12 cm column represented a reasonable balance between sensitivity and ion suppression, and thus was used for the following analyses unless specified (Figure 3A). This saturation point is expected to be proportional to the amount of resin in the column and may vary upon the properties of selected reverse-phase resins.

Second, we tested the effect of flow rate on capillary LC column performance (Figure 4). For 1 μg of loaded peptides, when the flow rate changed from 1.1 $\mu\text{l}/\text{min}$ to 0.15 $\mu\text{l}/\text{min}$, the best result was achieved with the flow rate between 0.15 $\mu\text{l}/\text{min}$ and 0.25 $\mu\text{l}/\text{min}$. A similar optimal flow rate of 0.25 $\mu\text{l}/\text{min}$ was found when 50 ng of peptides was loaded (Figure 4). This was not unexpected as slower flow rate resulted in more concentrated eluates and increase sensitivity. Further decrease of flow rate to 0.1 $\mu\text{l}/\text{min}$, however, worsened the results, which may be due to unstable electrospray. In our setting, the voltage was applied on a four-way tee for buffer splitting located ~ 20 cm away from the column tip³. The ionization was likely influenced by the flow rate. Furthermore, slower flow rates were associated to longer delays of elution (e.g. 20 min at 0.1 $\mu\text{l}/\text{min}$) because of dead volume (~ 2 μl). Therefore, we fixed the flow rate at 0.20 $\mu\text{l}/\text{min}$ for this 75 μm I.D. column in subsequent runs.

Third, we optimized the LC gradient range to fully utilize peptide identification power. As identifiable peptides were not equally distributed during the LC elution, it was desirable to expand the range within which most of the peptides were eluted. We performed a test run with 5–35% of buffer B in 45 min and found that 97% of the identified peptides eluted from 9%–30% of buffer B, equivalent to 13–32% of acetonitrile (Figure 5). Thus, we used the gradient range to 9%–30% of buffer B for this LC system.

Fourth, we adjusted elution time from 10 to 320 min to analyze 1 μg of peptides. The titration curve was not linear and started to plateau at the 160-min time point with 1,012 proteins identified. The 320-min run provided limited benefit with only 79 more proteins identified (Figure 6). We further tested different loading amounts (200 ng and 50 ng) and found the same plateau around 160-min gradient length (Figure 6). This phenomenon could be explained by the two effects with increased gradient length: (i) analysis time was longer to allow more MS/MS scans, and (ii) ion peaks may be broadened with less ion intensity. For example, when 50 ng of peptides was used in the series of analysis during 10 min to 320 min elution, the peak width of the TEF2 peptide (IGGIGTVPVGR) raised from 0.14 min to 0.90 min, whereas the peak height dropped from 100% to 16% in these runs (Figure 7A). The peak broadening caused by long gradient elution was also illustrated by global peak width distribution (Figure 7B). To this end, if sample amount was further limited, long elution time may be even detrimental to the analysis, because peptides signal may become too weak to be detected. To test this idea,

we carried out more analyses with 10 ng of loading amount (Figure 6). Indeed, the optimal elution time was decreased to 40 min. This evaluation is useful for selecting optimal gradient length based on the sample amount available.

Considering none of current LC/MS/MS system is capable of analyzing all peptides digested from real biological samples, the above optimization will facilitate protein identification as well as the analysis of posttranslational modifications (PTM). As sequencing coverage by MS/MS is critical for PTM analysis, we examined average sequencing coverage of proteins in one set of six runs (50 ng loading, 10–320 min). Like protein identification, the sequencing coverage is also increased with the elution time up to 80 min and then reached plateau (16.3% for 10 min, 23.7% for 20 min, 24.8% for 40 min, 24.9% for 80 min, 23.5% for 160 min and 21.0% for 320 min).

Conclusion

By systematical adjustment of parameters in shotgun proteomics, we optimized common parameters in our LC and MS settings on a 75 I.D. reverse-phase column. With the optimum flow rate set at 0.2 μ l/min and the gradient range set at 13–32% of acetonitrile, the gradient length should be adjusted according to the sample load amounts (e.g. 40 min for 10 ng of peptides, and 160 min for 1 μ g of peptides). Using the optimized settings, we were capable of identifying 1,012 proteins (clustered in 806 protein groups) from 1 μ g of tryptic yeast total cell lysate. Although some of the parameters may need adjustment when applied to different LC/MS/MS systems, the procedure and the data here are expected to be highly instructive for conducting efficient proteomics analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by NIH grants CA126222, AG025688 and NS055077 to J.P.

Abbreviations

LC	liquid chromatography
MS	mass spectrometer
SILAC	stable isotope labeling by amino acids in cell culture

References

1. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422(6928):198–207. [PubMed: 12634793]
2. Cravatt BF, Simon GM, Yates JR 3rd. The biological impact of mass-spectrometry-based proteomics. *Nature* 2007;450(7172):991–1000. [PubMed: 18075578]
3. Peng J, Gygi SP. Proteomics: the move to mixtures. *J Mass Spectrom* 2001;36(10):1083–91. [PubMed: 11747101]

4. Shen Y, Zhao R, Berger SJ, Anderson GA, Rodriguez N, Smith RD. High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Anal Chem* 2002;74(16):4235–49. [PubMed: 12199598]
5. MacNair JE, Lewis KC, Jorgenson JW. Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns. *Anal Chem* 1997;69(6):983–9. [PubMed: 9075400]
6. Hsieh S, Jorgenson JW. Preparation and evaluation of slurry-packed liquid chromatography microcolumns with inner diameters from 12 to 33 microns. *Anal Chem* 1996;68(7):1212–7. [PubMed: 8651494]
7. Motoyama A, Venable JD, Ruse CI, Yates JR 3rd. Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal Chem* 2006;78(14):5109–18. [PubMed: 16841936]
8. Vollmer M, Horth P, Nagele E. Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. *Anal Chem* 2004;76(17):5180–5. [PubMed: 15373459]
9. Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 2002;13(6):659–69. [PubMed: 12056566]
10. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 2000;72(6):1156–62. [PubMed: 10740853]
11. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, Horning S. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* 2006;78(7):2113–20. [PubMed: 16579588]
12. Makarov A, Denisov E, Lange O, Horning S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J Am Soc Mass Spectrom* 2006;17(7):977–82. [PubMed: 16750636]
13. Dephoure N, Zhou C, Villen J, Beausoleil SA, Bakalarski CE, Elledge SJ, Gygi SP. A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A* 2008;105(31):10762–7. [PubMed: 18669648]
14. Lu B, Motoyama A, Ruse C, Venable J, Yates JR 3rd. Improving protein identification sensitivity by combining MS and MS/MS information for shotgun proteomics using LTQ-Orbitrap high mass accuracy data. *Anal Chem* 2008;80(6):2018–25. [PubMed: 18275164]
15. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006;127(3):635–48. [PubMed: 17081983]
16. Seyfried NT, Xu P, Duong DM, Cheng D, Hanfelt J, Peng J. Systematic approach for validating the ubiquitinated proteome. *Anal Chem* 2008;80(11):4161–9. [PubMed: 18433149]
17. Spence J, Gali RR, Dittmar G, Sherman F, Karin M, Finley D. Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell* 2000;102(1):67–76. [PubMed: 10929714]
18. Xu P, Duong DM, Seyfried NT, Cheng D, Xie Y, Robert J, Rush J, Hochstrasser M, Finley D, Peng J. Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* 2009;(137):1–13.
19. Eng J, McCormack AL, Yates JR 3rd. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994;5:976–989.
20. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2003;2(1):43–50. [PubMed: 12643542]
21. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007;4(3):207–14. [PubMed: 17327847]
22. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC. Measurement of protein using bicinchoninic acid. *Anal Biochem* 1985;150(1):76–85. [PubMed: 3843705]
23. Tal M, Silberstein A, Nusser E. Why does Coomassie Brilliant Blue R interact differently with different proteins? A partial answer. *J Biol Chem* 1985;260(18):9976–80. [PubMed: 4019521]
24. Mann M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* 2006;7(12):952–8. [PubMed: 17139335]

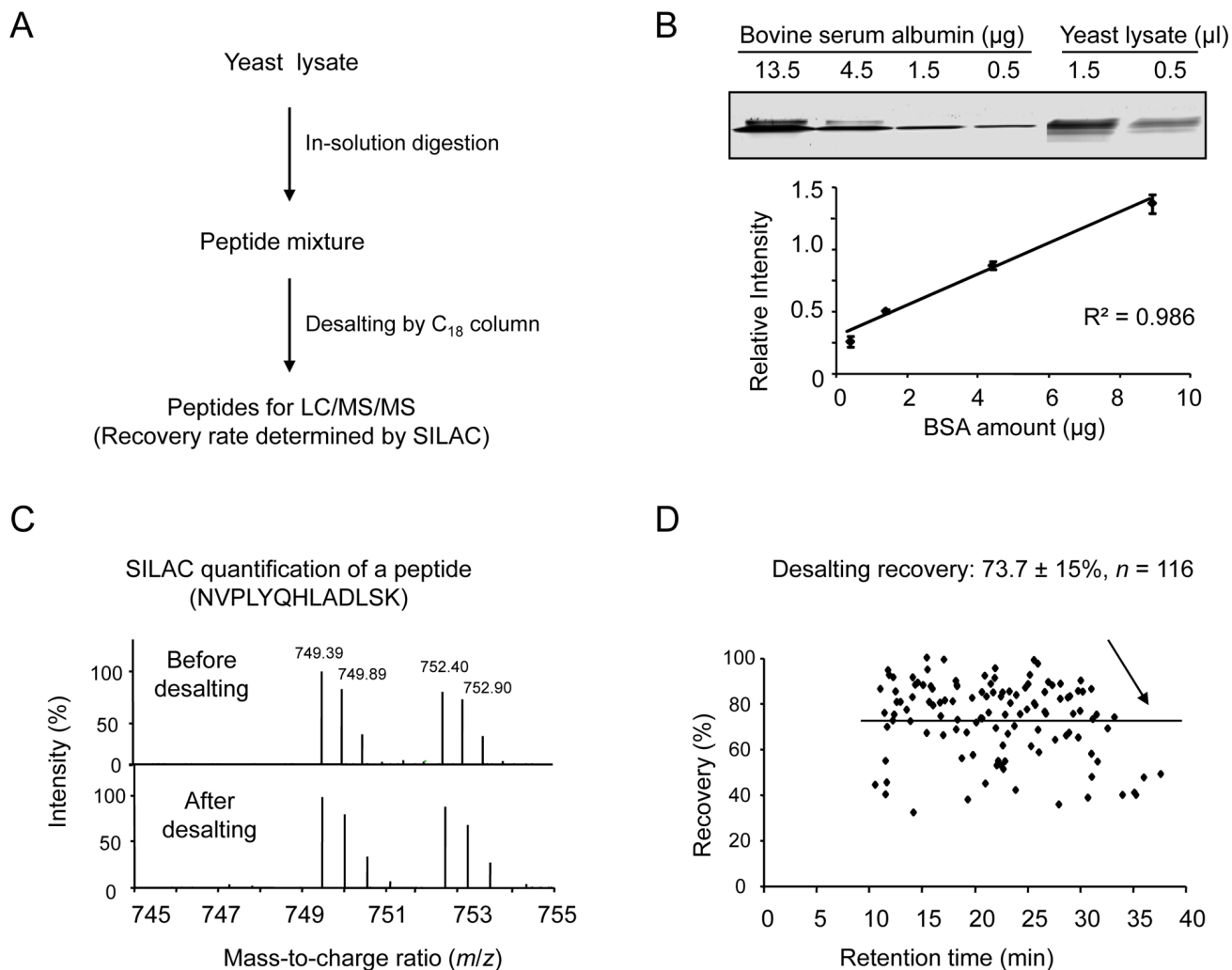


Figure 1.

Sample preparation for LC/MS/MS runs.

(A) Flow chart for quantifying protein and peptide used for optimization.

(B) Quantification of yeast lysate on a short 9% SDS gel with BSA as standard. The standard curve used for quantification was shown ($n = 3$).

(C) Peptide quantification by SILAC methodology. Same amount of peptides digested from heavy labeled yeast cells was spiked in the input and the elution fraction, and then analyzed by LC/MS/MS.

(D) Calculation of averaged peptide recovery from 116 peptides that are plotted according to retention time in LC/MS/MS. The value is shown as mean and standard deviation.

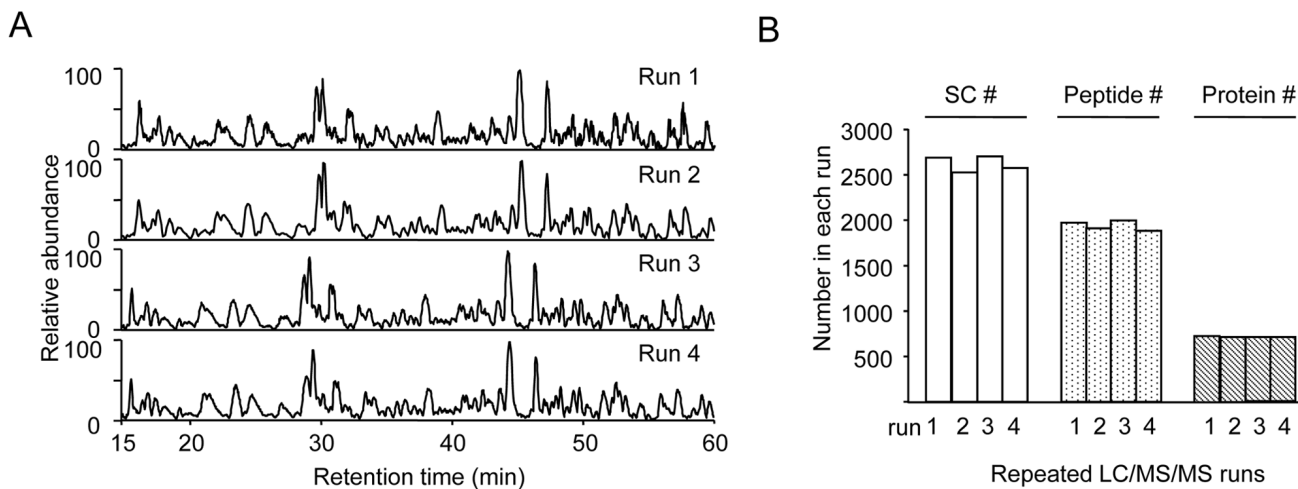
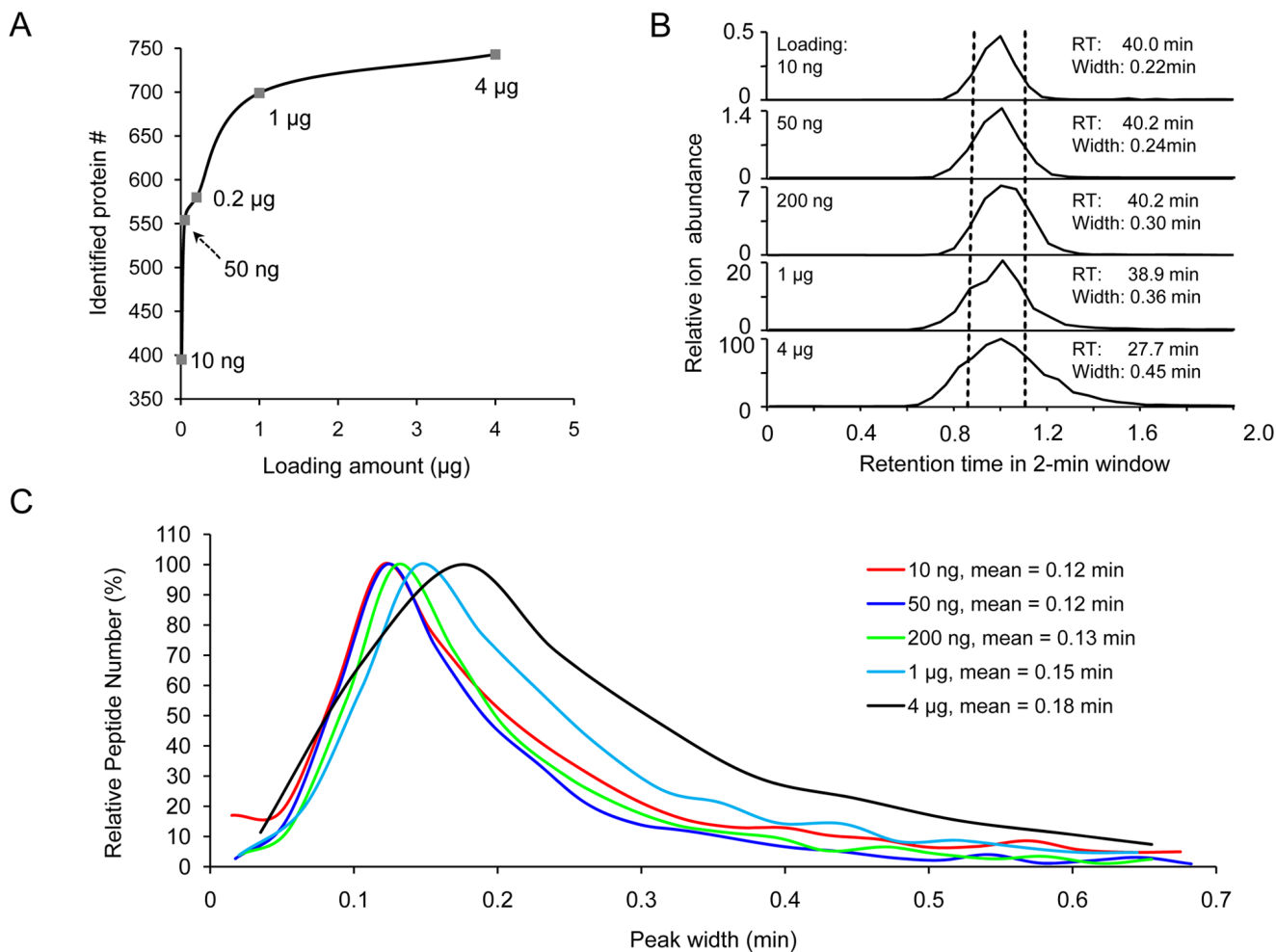


Figure 2.
 Evaluation of run-to-run variation of LC/MS/MS system.
 (A) Base-peak chromatogram of four repeat runs. Peptides derived from 1 μ g of yeast lysate were loaded by an autosampler during 15 min (not shown), and then eluted in a 5–35% gradient of buffer B over 45 min (15–60 min).
 (B) Comparison of accepted spectrum count (SC), peptides and proteins identified in four repeat runs.

**Figure 3.**

Optimization of loading amount for protein identification.

(A) Protein identification was increased with higher loading amount but could be saturated. The loading amount of peptides was calculated based on quantified protein concentration and peptide recovery. Peptides were eluted in a 5–35% gradient of buffer B over 40 min.

(B) Sample loading amount affects ion intensity and peak width in LC. The TEF2 peptide (IGGIGTVPVGR) of high abundance was shown as an example, including ion intensity, retention time, peak width at half height. The ion intensity in all runs was normalized to the intensity in the 4 μg run (100% = 4.7E8, mass range 512.9–514.6 m/z). (C) Distribution of peak width at half height for accepted peptides. For each run, the peak width data were first filtered by ion intensity (at least 2 signal-to-noise ratio), then grouped into ~20 bins for plotting histogram. The mean was derived by fitting most of the data points to Gaussian distribution.

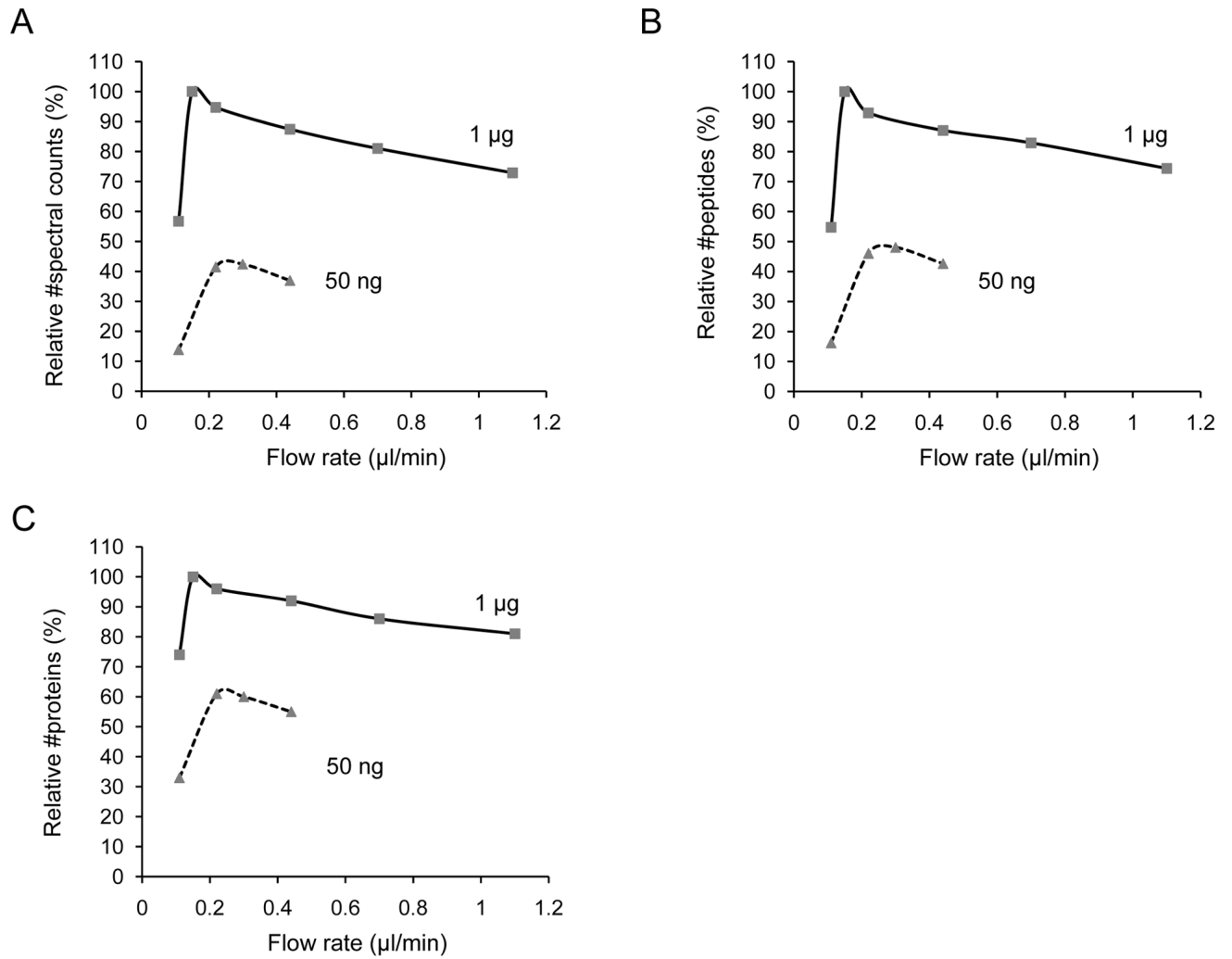


Figure 4. Optimization of flow rate in the LC/MS/MS runs to achieve maximum spectral counts (A), identified peptides (B) and proteins (C). Optimum flow rate for protein identification was ~0.2 µl/min, independent of the amount of sample loaded (1 µg: solid line; 50 ng: dashed line). The data point (0.22 µl/min, 1 µg loading) was repeated four times and led to almost identical result. The numbers shown in the figure were normalized to the maximal data point.

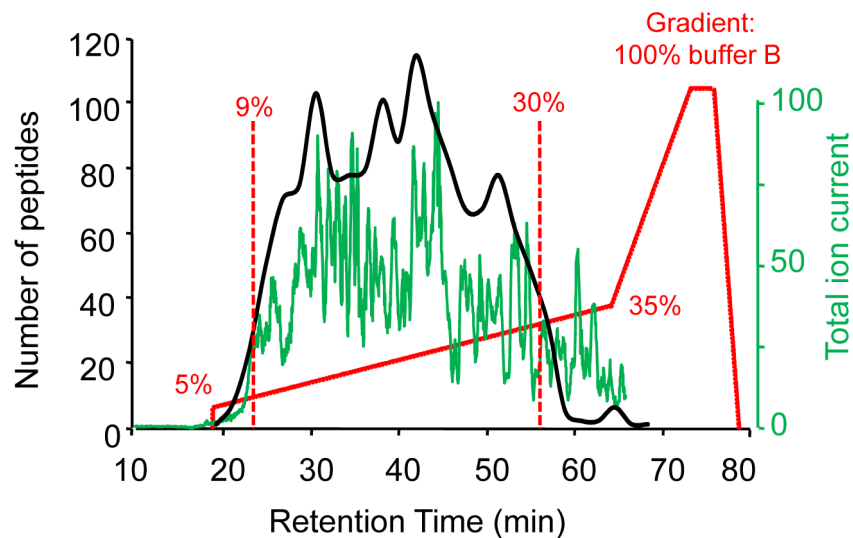


Figure 5.

Selection of gradient range for peptide elution. Peptides were loaded in the first 15 min, and eluted in a 5–35% gradient of buffer B at 0.2 $\mu\text{l}/\text{min}$ over 45 min. The curve of total ion current was shown in a dashed line and aligned with the gradient of buffer B after correction of dead volume. During the gradient, identified peptides were counted in every 2-min bin of elution and shown in a solid line. The optimized gradient range was 9–30% of buffer B that covered 97% of detected peptides.

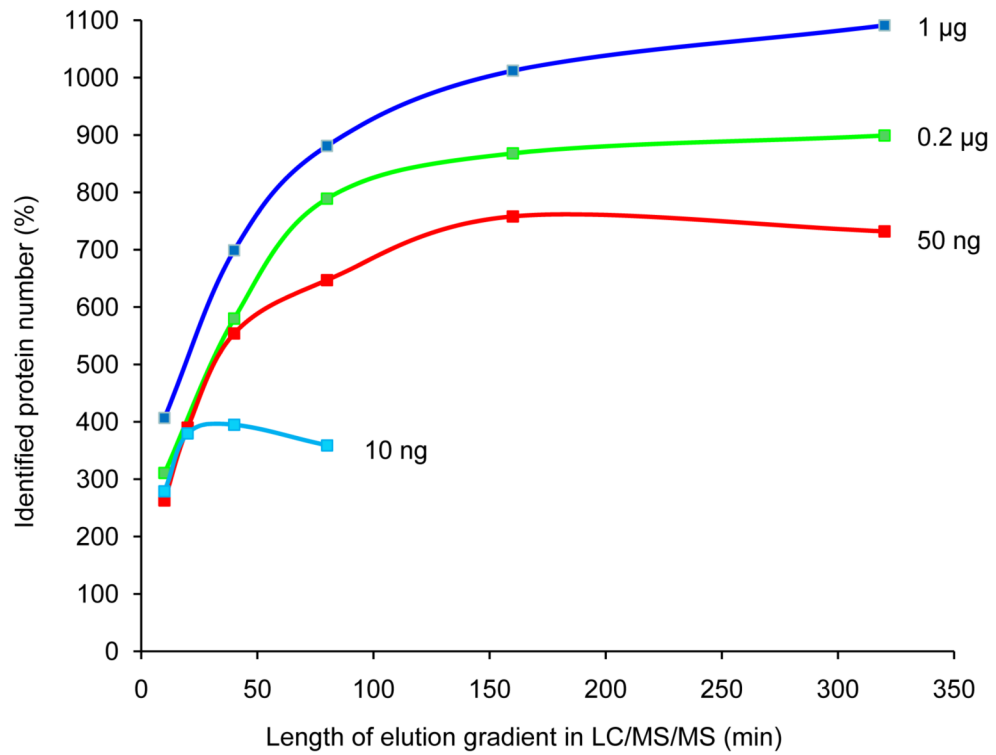
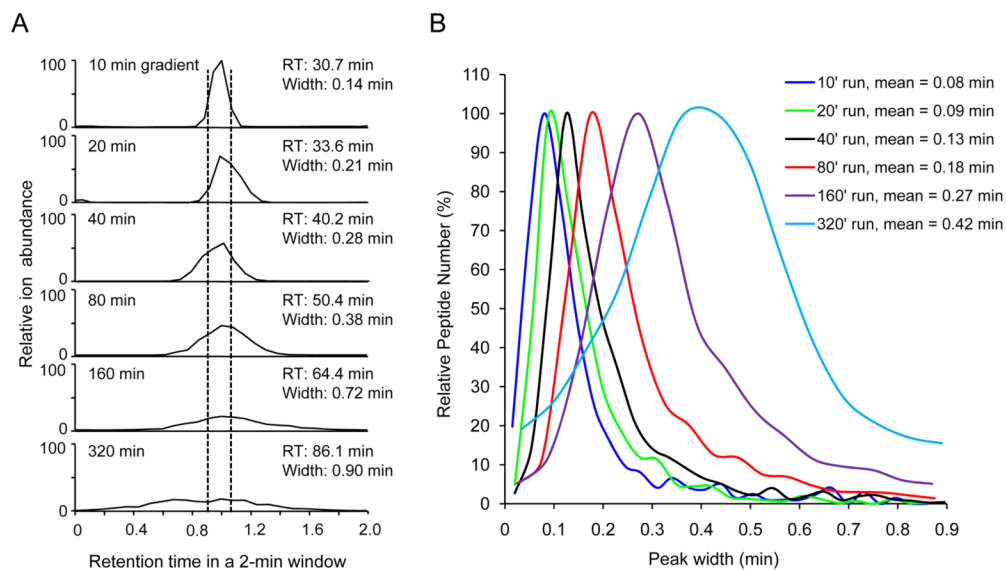


Figure 6. Characterization of sample loading amount and gradient length of LC/MS/MS for protein identification. The number of identified proteins varied upon the loading amount and the gradient length. Peptides were eluted in a 9–30% gradient of buffer B at 0.2 µl/min over 45 min.

**Figure 7.**

Analysis of peak width with the same loading but different gradient length.

(A) Comparison of peak width and intensity of one eluted peptide in LC/MS/MS runs (50 ng loading, 10–320 min gradient). The peptide ion (IGGIGTVPVGR in protein TEF2) of high abundance was extracted from the 10-min run, and aligned with the same peptide peak from the other selected runs. The ion intensities of all runs were normalized to the result of the 10-min run.

(B) Distribution of peak width at half height for accepted peptides in a set of runs (50 ng loading, 10–320 min gradient). The peak width data were filtered by ion intensity (at least 2 signal-to-noise ratio), then grouped into ~20 bins for plotting histogram. The mean was derived by fitting most of the data points to Gaussian distribution.

Table 1
Evaluation of false discoveries by the target-decoy strategy

LC-MS/MS Run	1	2	3
Pre-filtering			
Total SC ¹	6748	6860	6741
Target SC	4846	4814	4744
Decoy SC	1902	2046	1997
Estimated true matches ²	2944	2768	2747
Post-filtering			
Total SC	2713	2540	2652
Target SC	2703	2532	2642
Decoy SC	10	8	10
Total Proteins	659	687	609
Target Proteins ³	649	677	599
Decoy Proteins	10	10	10
Protein FDR(%) ⁴	3.0	2.9	3.3
Recovery of true matches ⁵	91.8%	91.5%	96.2%

¹ SC: spectrum count

² true matches = target SC - decoy SC

³ target proteins: matched by peptide(s) in the target database

⁴ False discovery rate (FDR) = (decoy proteins*2)/total proteins

⁵ Recovery rate = (post-filtering target SC)/Estimated true matches